

AFGHAN ROMANIZATION SYSTEM

4 May 2009

CONSONANT CHARACTERS

	Dari Letter	Pashto Letter	Unicode Code	BGN Translit	Ya'qūbī Translit	BGN Unicode	Sounds Like English	Examples
1	ب	ب	0628	b	b	0062	\b\ as in <i>bob</i>	بغلان / Baghlān
2	پ	پ	067E	p	p	0070	\p\ as in <i>pep</i>	پکتیا / Paktiyā
3	ت	ت	062A	t	t̤	0074	\t\ as in <i>tall</i>	تخار / Takhār
4		ټ	067C	t̤	t	1E6F	retroflexive \t̤\	کوټ / Kōṭ̤
5	ث	ث	062B	s̄	s̄	0073, 0304	\s\ as in <i>sand</i>	ورث / Waraṣ
6	ج	ج	062C	j	j	006A	\j\ as in <i>jazz</i>	جوزجان / Jowzjān
7	چ	چ	0686	ch	č	0063, 0068	\ch\ as in <i>chat</i>	چغچران / Chaghcharān
8	ح	ح	062D	h	ḥ	1E29	\h\ as in <i>hot</i>	حصارک / Hiṣārak
9	خ	خ	062E	kh	kh	006B, 0068	\ch\ as in German <i>ach</i>	خوست / Khōst
10		ځ	0681	dz	ṭ̤	0064, 007A	\dz\ as in <i>adze</i>	ځدران / Dzadrāṇ
11		څ	0685	ts	c	0074, 0073	\ts\ as in <i>gnats</i>	څمکنی / Tsamkanī

12	د	د	062F	d	ḍ	0064	\d\ as in <i>dog</i>	درزاب / Darzāb
13		ډ	0689	ḍ	d	1E0F	retroflexive \d\	ډنډ / Ḍaṇḍ
14	ذ	ذ	0630	z̄	z̄	007A, 0304	\z\ as in <i>zero</i>	گذرگاه نور / Gozargāh-e Nūr
15	ر	ر	0631	r	ṛ	0072	\r\ as in <i>run</i>	رشیدان / Rashīdān
16		ړ	0693	ṛ	r	1E5F	retroflexive \r\	کنړ / Kunaṛ
17	ز	ز	0632	z	z	007A	\z\ as in <i>zero</i>	زابل / Zābul
18	ژ	ژ	0698	zh	ž	007A, 0068	\zh\ as in <i>azure</i>	ژری / Zhaṛay
19		ږ	0696	<u>zh</u>	ž	007A, 0332 0068, 0332	retroflexive \zh\	زین تیږه / Zīn Tīz <u>h</u> ah
20	س	س	0633	s	s	0073	\s\ as in <i>sand</i>	سمنگان / Samangān
21	ش	ش	0634	sh	š	0073, 0068	\sh\ as in <i>ship</i>	شیرغان / Shibirghān
22		ښ	069A	<u>sh</u>	š	0073, 0332, 0068, 0332	retroflexive \sh\	پښتون کوټ / Pashtūn Kōt
23	ص	ص	0635	ṣ	s	015F	\s\ as in <i>sand</i>	صیاد / Ṣayād
24	ض	ض	0636	z̄	ḍz̄	1E95	\z\ as in <i>zero</i>	فیض آباد / Faīzābād
25	ط	ط	0637	ṭ	t̄	0163	\t\ as in <i>tall</i>	رباط سنگی / Rabāṭ-e Sangī
26	ظ	ظ	0638	ẓ	z̄	007A, 0327	\z\ as in <i>zero</i>	واغظ / Wāghaẓ

27	ع	ع	0639	‘	‘	2018	no sound	پل علم / Pul-e ‘Alam
28	غ	غ	063A	gh	gh	0067, 0068	no English equivalent / like \r\ in French <i>mer</i>	غزنی / Ghaznī
29	ف	ف	0641	f	f	0066	\f\ as in <i>fat</i>	فراه / Farāh
30	ق	ق	0642	q	q	0071	no English equivalent / closest sound is guttural \k\	قوش تپه / Qūsh Tepah
31	ک	ک	06A9	k	k	006B	\k\ as in <i>kitten</i>	کندهار / Kandahār
32	گ	گ	06AF	g	g	0067	\g\ as in <i>gain</i>	گومل / Gōmal
33	ل	ل	0644	l	l	006C	\l\ as in <i>land</i>	لغمان / Laghmān
34	م	م	0645	m	m	006D	\m\ as in <i>mint</i>	میمنه / Maīmanah
35	ن	ن	0646	n	n	006E	\n\ as in <i>night</i>	ننگرهار / Nangarhār
36		ڼ	06BC	ṇ	ṇ	1E49	retroflexive \n\	تنی / Taṇaī
37	و	و	0648	w	w	0077	\w\ as in <i>wait</i>	وردک / Wardak
38	ه	ه	0647	h	h	0068	\h\ as in <i>hot</i>	هرات / Herāt
39	ی	ی	0649	y	y	0079	\y\ as in <i>yard</i>	یمگان / Yamgān

VOWEL AND DIPHTHONG CHARACTERS

Sounds Like English	Dari Letter	Pashto Letter	Unicode Code	BGN Translit	Ya'qūbī Translit	BGN Unicode	Examples
40 short \a\ as in <i>fat, man</i>	unwritten		---	a	a or â	0061	غزنی / Ghaznī
41 short \i\ or \e\ as in <i>sit</i> or <i>ten</i>	unwritten		---	i or e ¹	i or e	0065/0069	هرات / Herāt
42 short \u\ or \o\ as in <i>put</i> or <i>no</i>	unwritten		---	u or o ²	o or u	006F/0075	ارزگان / Uruzgān
43 Pashto vowel \ə\ as in <i>up, jug</i>	unwritten		---	ê	ə or ě	00EA	برمل / Bêrmal
44 long \aa\ as in <i>car, fall</i>	آ	آ ³	0622	ā	ā	0101	آب بند / Āb Band
45 long \o\ as in <i>boat, toe</i>	و	و	0648	ō	ō	014D	لوگر / Lōgar
46 long \oo\ as in <i>moon, rule</i>	و	و	0648	ū	ū or u	016B	نورستان / Nūristān
47 diph. \ow\ as in <i>now, town</i>	و	و	0648	ow	aw or āw	006F, 0077	ده نو / Deh-e Now
48 long \ee\ as in <i>meet, see</i>	ی	ی	06CC	ī	i or ī	012B	بادغیس / Bādghīs
49 diph. \ay\ as in <i>date, maid</i>	ی	ی	06CC	ay or ē ⁴	ay or ē	0061, 0079 or 0113	زیباک / Zaybāk
50 long \aa\ as in <i>car, fall</i>	ی	ی	0649	á	ā	00E1	موسی قلعه / Mūsá Qal'ah
51 Pashto vowel \ā\ as in <i>rain, tail</i>	---	ې	06D0	ē or ay	ē	0113	رېگستان / Rēgistān
52 Pashto vowel \əy\ word ending	---	ی	06CD	êy	ə y or ay	00EA, 0079	څوکی / Tsowkêy

53	diph. \ī\ as in <i>my, eye</i>	ی	ی	06CC	āī	ay or āy	0061, 012B	میدان شهر / Maīdān Shahr
54	Dari <i>izāfat</i> \eh\ or \yeh\ as in <i>day</i> or <i>yea</i>	unwritten	---	---	-e / -ye	-i- or e		سر پل / Sar-e Pul قلعه کاه / Qal‘ah-ye Kāh
				OR				

OTHER DIACRITICAL MARKS AND LANGUAGE CONVENTIONS

Sounds Like	Dari	Pashto	Unicode	BGN	Ya'qūbī	BGN		
English	Letter	Letter	Code	Translit	Translit	Unicode	Examples	
55 \ī\ as in <i>my, eye</i>		ای	ای	0627, 06CC	āy ⁵	ay or āy	0101, 012B	دایکندی / Dāykundī
56 \īa\ as in <i>Alexandria</i>		یا	یا	06CC, 0627	tā or iyā ⁶	ya or yā	012B, 0061	غوریان / Ghōrīān
				OR				گیان / Giyān
57 (no sound)	ا ⁷	ا ⁷	0627	---	---	---		ارزگان / Uruzgān
58 (no sound)	و	و	0648	w	w	0077		خواجه دو کوه / Khwājah Dō Kōh
59 (glottal stop)	ئ	ئ	0626	'		2019		درواز پائین / Darwāz-e Pā'īn
60 \heh\ as in <i>hay</i>	ه	---	06C0	. . .h-e		---		کوه صافی / Kōh-e Ṣāfī
61 (no sound)	ـ	ـ	0651		(consonant doubler)	---		محمد آغه / Muḥammad Āghah
62 -----	الله	الله	FDF2	Allāh	?	---		عبدالله شيله / ‘Abdullāh Shēlah

FOOTNOTES

- 1 (Line 41) Both *e* and *i* are available to romanize this short vowel, depending on local usage and/or root language. In cases where the sound is uncertain, *i* is the default romanization in BGN/PCGN standardization procedures.
- 2 (Line 42) Both *o* and *u* are available to romanize this short vowel, depending on local usage and/or root language. In cases where the sound is uncertain, *u* is the default romanization in BGN/PCGN standardization procedures.
- 3 (Line 44) An *alif* with *mad* (ﺍَ) is written only in the initial position by BGN/PCGN standardization procedures, in keeping with Persian language family standards of use of the Arabic alphabet. The same letter written in a medial or final position is written ﺍ . .
- 4 (Line 49) The preferred romanization for this sound in Dari words and/or in Dari-speaking areas is *ay*, while the same sound is romanized *ē* in Pashto words and/or in Pashto-speaking areas.
- 5 (Line 54) Afghan scholarly sources handle this digraph as a vowel/consonant combination, not as two vowels or as a diphthong, although its pronunciation is identical, or nearly so, to the diphthong sound which does not use an *alif* in Arabic letters.
- 6 (Line 55) Per BGN/PCGN standardization procedures, the preferred romanization of this letter combination in Dari words and/or in Dari-speaking areas is *īā*, while *iyā* is preferred for Pashto words and/or in Pashto-speaking areas.
- 7 (Line 57) This convention applies only to the initial *alif* which has no sound of its own, serving as a seat for an unwritten short vowel. This is in keeping with Persian language family standards of use of the Arabic alphabet whereby words that begin or end with an unwritten short vowel must utilize a silent consonant to carry the sound in written form.

NOTES

- 1 Unicode values listed in this table are required for all BGN/PCGN production, to ensure standardization and to minimize confusion from competing representations of a given letter. Notably, the Persian ك (Unicode value 06A9) is prescribed rather than the Arabic ك (Unicode value 0643 or FEDA or FED9), the Persian گ (Unicode value 06AF) is prescribed rather than ك (Unicode value 06AB) or گ (Unicode value 06B0) or ك (Unicode value 0643 or FEDA or FED9), and the Pashto letter خ (Unicode value 0681) is prescribed rather than the *heh* with a dot above and a dot below (Unicode value unknown). For the letter ی in its many variations, care must be exercised to follow this romanization guide's prescriptions to eliminate confusion for search engines and software. BGN/PCGN does not use the letters ي (Unicode values 064A or FEF1) or ی (Unicode value FEEF) in any Afghan word.
- 2 Occasionally the letter sequences كه, زه, سه, and گه occur without intervening vowels. They may be romanized *k·h*, *z·h*, *s·h*, and *g·h* in order to differentiate these romanizations from the digraphs *kh*, *zh*, *sh*, and *gh*, which are used to represent the letters خ, ژ, ش, and غ. Additionally, the Pashto letters خ and دز, routinely romanized *ts* and *dz*, may be alternatively romanized *tš* and *dž* when for special reasons it is desired that confusion be avoided with the character sequences تس (*ts*) and دن (*dz*), respectively.

- 3 The vagaries of written Afghan languages, as pertains to spacing and word division, are addressed as follows. Spaces may be added to or subtracted from Afghan words written in Arabic script, for the purposes of standardization. This is particularly relevant when the words are hand-written, are rendered “artistically”, or express other such non-standard flourishes, as long as the sense of the toponym, word, or phrase is not compromised. Analysts proceed in this regard conservatively.

Romanized toponyms are typically divided into constituent words (spaces and other grammatical rules applied) when those words can stand independently, for purposes of standardization and minimization of confusion, particularly in situations where Afghan writers are inconsistent in their application of spacing and word breaks. When the Afghan word or suffix is only used in combination with other nouns or adjectives, then it should be appended to the preceding word in its romanization. This includes (but is not limited to) *-ābād*, *-zaī*, *-zādah*, *-tū*, *-wand*, *-gaī*, *-kaī*, *-pūr*, *-dēsh*, *-lar*, *-lī*, *-lū* and *ullāh*, as, for example, seen in *Raḥmatābād* (رحمت آباد) and *Raḥmatullāh* (رحمت الله), but *Raḥmat Khēl* (رحمتخيل) and *Raḥmat Shahr* (رحمتشهر).

- 4 The one-letter words د (Pashto) and و (Dari) are romanized *dê* and *wa*, respectively.
- 5 The word الله, meaning God, should always be romanized *Allāh*, except as specified in note 3. Note that the Unicode value FDF2 spells *Allāh*, but omits the *alif* in some common fonts, including Times New Roman. If in doubt, try in Arial Unicode MS to verify.
- 6 In names of Arabic origin, the *l* of the definite article *al* is assimilated before the ‘sun letters’ *t*, *ṣ*, *d*, *ḏ*, *r*, *z*, *s*, *sh*, *ṣ*, *ẓ*, *ṭ*, *ẓ*, and *n*. In its romanization, the article should be separated from the name it precedes and should not be capitalized except at the beginning of a name, e.g. جبل السراج → *Jabal as Sarāj*
- 7 Specifically Arabic names, particularly those of Islamic significance, often include a consonant which would ordinarily be written with a *tashdīd* (*shaddah*) in Arabic, which may be romanized by writing the equivalent consonant twice, e.g., مُحَمَّد → *Muḥammad*. This rule is particularly in force if the word is commonly romanized with a doubled consonant by Afghan naming authorities. For example, *ḥājī* is rarely seen in Afghan-produced romanizations with a doubled consonant, but *‘Abbās* is. The doubled *y* consonant is almost always seen thusly, as in *Sayyid* or *Qayyūm*.
- 8 In Afghan names which contain an *izāfat*, it should be romanized as *-e* or *-ye* according to common pronunciations, but generally, *-e* is used if the preceding word ends with a consonant other than silent *heh*, and *-ye* if the preceding word ends with a vowel sound. Scholarly sources indicate that *heh* is silent in *darah* and *qal’ah* (thus *darah-ye*, *qal’ah-ye*), but lightly spoken in *kōh* and *chāh* (thus *kōh-e*, *chāh-e*).

- 9 The character sequence **خو**, when followed by **١** or **ى** should be Romanized *khwā* or *khwī*, when the *w* is not pronounced, or only weakly so, as in *khwājah*.
- 10 Plural nouns ending in *-hā* or *-ān* should always be romanized as a single word, regardless of whether a space appears in an Arabic script source.
- 11 Although Arabic script is written from right to left, numerical expressions, e.g. **١٩٦٨** → 1968 are written from left to right.

٠	١	٢	٣	٤, ٤	٥, ٥	٦, ٦	٧	٨	٩
0	1	2	3	4	5	6	7	8	9